A guide to...

...Survey Design & Analysis
This guide gives practical advice and guidance to people who need to design surveys and analyse the resulting data. You may also want to look at:

- Questionnaire Design
- Invitation to Tender Template for Surveys/Research

We have produced a series of guides which covers four broad areas:

- Surveys
- Performance management
- Collecting and using data
- Statistical processes

This series of guides is available electronically on our website, www.dataunitwales.gov.uk/guides

We also provide advice, guidance and training support relating to each of the documents in the series. For further information contact:

Local Government Data Unit ~ Wales
029 2090 9500

or visit our website: www.dataunitwales.gov.uk


First published 2005
Second edition 2009

Published by:
Local Government Data Unit ~ Wales
3 – 7 Columbus Walk, Cardiff, CF10 4SD
Telephone 029 2090 9500
Email enquiriess@dataunitwales.gov.uk
Web www.dataunitwales.gov.uk

© Local Government Data Unit ~ Wales, 2009 (Data Unit)
This material can be reproduced as long as the Data Unit is acknowledged as the original author. Any queries should be sent to be the Local Government Data Unit ~ Wales at the above address or via e-mail: enquiries@dataunitwales.gov.uk
## Contents

1. **Introduction**
   - 1

2. **What is a survey?**
   - 1
   - 2.1 Why survey?  1

3. **Preparation**
   - 2
   - 3.1 Before doing a survey  2
   - 3.2 Types of survey  3

4. **Sampling**
   - 4
   - 4.1 Why do we need a sample?  4
   - 4.2 Sampling frames  5
   - 4.3 Sampling error  6
   - 4.4 Size of the sample  6

5. **Questionnaire design**
   - 7
   - 5.1 Introducing the questionnaire  7
   - 5.2 The Data Protection Act  8
   - 5.3 Structure and layout  8
   - 5.4 Questions  9
   - 5.5 Coding  11
   - 5.6 Routing  12

6. **Piloting the survey**
   - 12

7. **Analysing the results**
   - 14
   - 7.1 Data entry  14
   - 7.2 Analysis  15

8. **Reporting**
   - 17
   - 8.1 Presenting the results  17

**Appendix 1**: Example questionnaire  18

**Appendix 2**: Additional support  21

**Appendix 3**: Further reading  22
1. Introduction

This guide focuses on good practice relating to survey design and analysing the data gained through using surveys. We have included references within this document to information which will support larger or more detailed work.

This guidance is aimed at local authority officers who design and use questionnaires to show continuous improvement in their service areas and who, often, have little or no training in this field.

2. What is a survey?

Using a survey is a way of collecting information. Local authorities constantly use surveys, whether to measure public opinion through satisfaction surveys or housing conditions through more service specific surveys. To extract the information needed from a survey, the process must be planned effectively and the survey correctly designed and implemented.

This guide gives you basic advice on the key points that you should consider when designing and analysing survey results. References are given in Appendix 3 that will direct you to more detailed guidance.

2.1 Why survey?

The aim of a survey is to collect data that is accurate, precise and valid.

Accuracy: the data collected should be a good reflection of the variables being measured.

Precision: the quality of the data depends on how the survey has been designed and how questions are worded. A measure of the quality is the precision – how close to the real values will those from the survey be?

Valid: does the data measure what you set out to measure?
3. Preparation

3.1 Before doing a survey

You will need to do some groundwork before you undertake a survey.

You should ask:

**What are the survey’s objectives?**
The main question is *what information is required from the survey?* Define exactly what you want from the survey and list specific questions that you would like the survey to answer. This includes any combination of answers: for example, if the survey collected the respondent’s age in 10 year groups, it would not be possible to define a new variable ‘Pensionable age’ because there would not be a break at 64.

**Is a survey the best way to get the data?**
Could this information be extracted by analysing data that already exists? Would a focus group be a better way of getting hold of your data? (These are better for getting *qualitative data;* thoughts and opinions about alternatives or proposals.)

**Is it possible to get the information that you need?**
You can get hold of most information but it may not be readily available or may not be measured easily. For example, you may want to measure the amount of calcium that a group of children consume, but this is not practically achievable since it is not eaten as a separate item but included in many others in invisible form. Similarly, if you were to ask a group of youths about their anti-social behaviour, you may not get truthful answers; or if you asked alcohol consumers about how much they drank in the last month, again this is notoriously under-reported.

**Is the survey worthwhile?**
Surveys can be expensive. Find out whether the survey has been done before, so that you can avoid duplication and unnecessary costs. Perhaps it’s been done by another authority. Are there any alternatives or administrative data sources that could be useful? Is it possible to gather more than one information requirement to share the overheads?

**Literature review?**
This can be useful to check whether similar surveys have been carried out. These may be a source of comparative data and could be useful to get tested questions. They can also be used as an indication of what not to do in a survey. It may be possible to contact the people who worked on these surveys and discuss problems and issues with them.

**How do I do it?**
Once you have decided to carry out the survey and your objectives are clear, you will need to decide on the type of survey you’d like to undertake.
## 3.2 Types of survey

There are various types of survey. Which one you choose will depend on the survey’s purpose, the timescale for collecting the results and the resources available. The types of survey that are commonly used for collecting quantifiable data in a structured way, along with the advantages and disadvantages of using that type of survey, are summarised below.

<table>
<thead>
<tr>
<th>Survey Type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Postal surveys</strong></td>
<td>• Wide distribution&lt;br&gt;• Low cost&lt;br&gt;• No interviewer bias&lt;br&gt;• Easier to conduct in-house&lt;br&gt;• No embarrassment for respondent in giving answers to another person</td>
<td>• Low response rate&lt;br&gt;• Not suitable for some subgroups e.g. low literacy&lt;br&gt;• Questionnaire: must be simple and short, not too probing</td>
</tr>
<tr>
<td><strong>Face-to-face</strong></td>
<td>• Higher response rates&lt;br&gt;• Best method for asking open and closed questions&lt;br&gt;• Flexibility (able to question respondent to get a full reply)</td>
<td>• High cost&lt;br&gt;• Time consuming&lt;br&gt;• Possible interviewer bias</td>
</tr>
<tr>
<td><strong>Telephone interviews</strong></td>
<td>• Low cost&lt;br&gt;• Quick&lt;br&gt;• Reasonable response rate</td>
<td>• Potential bias: i.e. Individuals with no landline phone or ex-directory&lt;br&gt;• Questionnaire constraints: Limited to a few response categories; unsuitable for sensitive questions</td>
</tr>
<tr>
<td><strong>Focus groups</strong></td>
<td>• Low cost&lt;br&gt;• In depth responses to questions&lt;br&gt;• Detailed probing for reasons&lt;br&gt;• Informal - with more relaxed respondents</td>
<td>• Possibly too much information gathered - needs analysis&lt;br&gt;• Need good planning and a good facilitator</td>
</tr>
<tr>
<td><strong>Internet surveys</strong></td>
<td>• Data automatically collected as soon as respondent presses ‘submit’ on computer&lt;br&gt;• Low cost&lt;br&gt;• No interviewer bias&lt;br&gt;• Respondents are automatically ‘routed’ through the survey</td>
<td>• Potential bias: i.e. individuals with no internet access&lt;br&gt;• Low response rate&lt;br&gt;• Sample sometimes self-selecting: i.e. survey ‘pop ups’ when an individual is on a website asking them to fill it in</td>
</tr>
</tbody>
</table>
4. Sampling

4.1 Why do we need a sample?
It is usually not possible to capture data on a whole population; the survey that comes closest to sampling a whole country is the Census - but even that does not collect data on 100% of the population. To examine a sample of a population, there are two basic forms of sampling.

**Probability sampling** - selecting a random sample.

**Non-probability sampling** - selecting a non-random sample.

**Probability sampling** is a method that allows every individual or household in a population to have an equal chance of being selected. The individuals or households are selected before the survey takes place and cannot be changed.

**Example 1**
The Welsh Health Survey is carried out with the aim of collecting information on a cross-section of the population, with a range of illnesses or disabilities, and on similar groups of healthy people. The survey’s sample was taken from the Postcode Address File.

- Welsh Health Survey, Welsh Assembly Government

The **non-probability sampling** method does not have a random element. The questionnaire is aimed at a group of individuals or objects (e.g. houses, cars) that would not be practical or possible to obtain at random.

**Example 2**
A street interview, or a self-selected sample are examples of non-probability sampling - such as surveying two groups of children within a school to compare teaching methods. However, we cannot be certain that the samples are truly random and you should take care when interpreting the results. Some web sites have surveys that ‘pop up’ on screen when you visit the web site – the samples are self-selecting and almost certainly not representative.
### Probability vs non-probability sampling

<table>
<thead>
<tr>
<th></th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Probability sampling</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Simple random sampling | • Can be generalised to the whole group from which the sample is selected  
                         | • Better quality estimates                             | • Costly                                               |
| Stratified sampling   | • More accurate results                                  | • Time consuming                                      |
|                      |                                                          | • Requires a sampling frame                           |                                                      |
| **Non-probability sampling** |                                                          |                                                      |
| Self-selected samples | • Higher response rates                                 | • Cannot be generalised to the population             |
| Street interviews     | • Best method for asking open and closed questions       | • Poor quality estimates                              |
| Telephone sample      | • Flexibility (able to question the respondent to get a full reply) | • Cannot tell if results are accurate                  |
| Judgmental sample     |                                                          |                                                      |

#### 4.2 Sampling frames

Before selecting a sample from a certain population you need to establish a list of the units within that population. For example, in a local authority area’s population, a unit can consist of an individual person or an individual household. The list must always be structured so that all the items have an equal chance of being selected.

A sample is then taken from a list of units. This is called a sampling frame.

A sampling frame ensures that you represent every aspect of the population that you are targeting. It is easy to get a sampling frame and draw samples from it.

Many possible sampling frames exist within a local authority. Some that may already exist are Council Tax databases, social services databases, electoral registers, and Local Land and Property Gazeteers although these may relate to specific groups. It is necessary to ensure that the lists are fully representative of the population that you are going to sample.
4.3 Sampling error

With any survey, there are errors associated with the results and you should consider this when you are interpreting them. If your figures are based on a sample from the population, as opposed to the whole population, the result will be subject to sampling error. The smaller the sample size, the greater the sampling error. You can measure the sampling error by calculating a confidence interval for the estimate. A 95% confidence interval is an interval within which we are 95% sure that the true value of the proportion for the population lies.

If you wanted to calculate a 95% confidence interval for a proportion of the total sample n (where p is the sample estimate), you should apply the following formula:

\[ p \pm 1.96 \sqrt{\frac{p(1-p)}{n}} \]

1.96 relates to the 95% confidence interval. For a 99% confidence interval, 1.96 would become 2.57. These figures are taken from published statistical tables.

Example 3

If 24% of people would actually complain about their bus service, but we did not know this, we could undertake a survey to get an estimate.

A survey of 500 people was carried out, and from the results, we found that 25% complained about their local bus service.

A percentage is just a different way of expressing a proportion (proportion = percentage/100) so in this case, \( p = 0.25 \) and \( n = 500 \).

A 95% confidence interval for \( p \) is

\[ 0.25 \pm 1.96 \sqrt{\frac{0.25(0.75)}{500}} = (0.21, 0.29) \]

So, we can be 95% confident that the true percentage of the population that would complain about their bus service is between 21% and 29%. The true value is within this range.

4.4 Size of the sample

Generally, the greater the sample size, the smaller the sampling error, and the estimates will be more accurate. We can illustrate this by taking Example 3 and increasing the sample size (n) to 2,500; the sampling error will decrease, and so reduce the size of the confidence interval, as shown in Example 4.

Example 4

A survey of 2,500 people was carried out and from the results; we found that 25% complained about their local bus service. In this case, \( p = 0.25 \) and \( n = 2,500 \).

A 95% confidence interval for \( p \) is \( (0.23, 0.27) \).

So, we can be 95% confident that the true percentage of the population that would complain about their bus service is between 23% and 27%.
Because the square root is present in the sampling error formula, doubling the sample size does not half the sampling error, but reduces it by a factor of the square root of 2, (i.e. 1.414). Quadrupling the sample, however, does halve the sampling error (as the square root of 4 is 2!).

When choosing a sample size, we have to consider the expected response rate because the error calculations are based on the achieved sample. For example, if we decide to sample 500 respondents for a survey and we expect that the response rate will be about 40%, we’ll actually achieve a sample of 200. If this produces a sampling error that is too large then you will need to adjust the initial sample size to reflect the low expected response. One of the objectives of a pilot survey (explained fully later) is to test the response rate.

Key points
• The sample should be properly selected from the population.
• The sample should be sufficiently large so that the sampling error is minimised and the calculated confidence interval is judged sufficiently narrow (e.g. an error of plus or minus 10% of the estimate would generally be considered acceptable for social data).

5. Questionnaire design

To get data that meets the study’s objectives, the questionnaire’s design is extremely important. The design can affect the number of responses (response rate), the data’s quality, and also how the data will be handled once it’s returned. For more detailed information, see the Data Unit’s guide on ‘Questionnaire Design’.

5.1 Introducing the questionnaire

When introducing the questionnaire to the respondents (either by post in an introductory letter, by telephone, on-line or face-to-face), it is important to note all of the following points:

**Interviewer surveys:**
• the organisation responsible for the research;
• the research’s purpose and its importance;
• the importance of the respondent’s participation; and
• emphasise confidentiality.

**In addition, for postal surveys:**
• the importance of reply i.e. Your response will allow us to....;
• an explanation of how the respondent was selected;
• offer a report on the results - or an indication of where and when the results will be available;
• thank the respondent for their time;
• personalise it: sign and date;
• enclose a stamped self addressed envelope and the date it needs to be returned by;
• use stamps as opposed to franking the envelopes as this has been proved to increase response; and
• be brief.
5.2 The Data Protection Act

You may also have to give some additional information to respondents under the Data Protection Act, regardless of your collection method. It is essential that your respondents know whether the data will display personal attributes (age, sex, ethnicity etc.) that can be traced back to them.

If it is possible to trace the information back to the respondent, you must tell the respondent at the time of the survey:

- what will happen to the data;
- how it will be used; and
- who will be the ultimate data custodian.

In technical language, the respondent has to know who the **Data Controller** is and has to be given a **Fair Processing notice** (explaining who will have access to the data and for what purpose).

This usually takes the form of an information leaflet for an interviewer-conducted survey. Similar information has to be given even if the data is anonymised before the Data Controller receives the information (for example, where the respondent’s name and address are removed from the record by the research company before the data is passed to the ultimate user).

5.3 Structure and layout

The questionnaire must be easily understood if you are to get the response that you need. Here are some key points to consider regarding the way that the questions are presented on paper.

- Well spaced - include white space and clear instructions for the respondent throughout the survey/questionnaire.

- Make sure that the question and answer formatting allows the respondent to fill in the questionnaire easily - allow space for handwriting; usually double line space.

- Instructions for completing each question should be clear - tick relevant box; circle the corresponding number; enter the number of people in your household; write the postcode of your place of work.

- Questions should be numbered and not split between pages.

- Routing: if some questions are not to be answered by all, instructions should be clear as to which questions are answered next - such as which question needs answering following an earlier response. For example, if Yes, please answer Question 4, otherwise go to Question 6.

- The front page design is important. You can get a higher response rate from a well designed front page.

**What is routing?**

Some questions may only be applicable to people who give certain answers in other questions. For example, if a respondent answered that they had no children they would not need to answer a section of questions about the age and sex of the children. This can be avoided by routing the respondent to the next set of applicable questions. This will avoid receiving incorrect or irrelevant information.
Survey Design & Analysis

Question order

The way the questions are ordered is very important. If the respondent feels that it is difficult to complete the questionnaire then they may refuse to finish it. By putting more general information at the beginning of your questionnaire you build up a relationship that encourages the respondent to continue. It may be better to leave questions that ask for sensitive information, such as financial details, to the end of the questionnaire. By doing this, even if the respondent refuses to complete the questionnaire, you will have collected the majority of your information already.

- First questions should be easy, short, interesting and applicable to every respondent.
- Sensitive questions, such as those on income, should be placed later in the questionnaire.
- Questions on the same subject should be grouped together.
- Demographic questions, such as those on age, gender and people’s relationship within the household should be placed together.

5.4 Questions

The quality of the data that you get will depend on the quality of your questions. Unclear questions will get unclear answers – which are then unusable!

Some points to note:

- Do not use technical terms; use everyday language.
- Write clear and focused questions; no ambiguity.
- Do not ask more than one thing in a question, e.g. How satisfactory is the refuse collection and cleanliness in your area? It will be hard to know whether the response relates to refuse collection or cleanliness.
- Do not re-invent questions; question banks are valuable tools which identify possible questions you could include in a questionnaire. See Appendix 2 for information on question banks and harmonised questions.
- Avoid leading questions e.g. Many people believe that regular exercise is important to health. Do you do any regular exercise?
- Avoid questions involving memory – very few people can accurately identify events precisely in time unless they are significant. So a question such as How many times have you visited the cinema in the last twelve months? will be very difficult for the people who go reasonably frequently, but be very easy for those who didn’t go at all!
- Understand that honesty is not always forthcoming in answers: sometimes because the respondent wants to delude themselves and sometimes because they do not want to admit the truth; answers to questions relating to alcohol consumption and smoking are notorious for this.
There are two distinct types of questions in a questionnaire:

**open-ended** and **closed**.

### Open-ended questions

Respondents answer the questions in their own words.

- **Completely open-ended** - An exploratory question with a wide range of answers and so is difficult to process, e.g. What are your views on the catering at the venue?
- **Directed open-ended** - A more focused question, e.g. Where is the venue located?
- **Restricted open-ended** - Very focused and usually a number, e.g. How many children do you have?

### Closed questions

Respondents are able to choose from a list of options. We suggest that you include an **Other - please specify** category in multi-choice questions.

- **Dichotomous** - Yes/No (can include a *Don’t know* and/or a *Not applicable* category).
- **Multi-choice** - Offer a mutually exclusive and exhaustive list of choices (i.e. the items in the list should not overlap and should cover every possibility). This type of question sometimes allows the respondent to give more than one answer to the question. Example 5 shows how not to list choices.

### Example 5

**How far is your home from the nearest bus stop?**  

*Please circle one number only*

<table>
<thead>
<tr>
<th>Distance</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 1 mile</td>
<td>1</td>
</tr>
<tr>
<td>1-5 miles</td>
<td>2</td>
</tr>
<tr>
<td>5 miles or more</td>
<td>3</td>
</tr>
</tbody>
</table>

The problem with this list is that if the respondent lived exactly 5 miles away, which number should they circle - 2 or 3? To rectify this, Option 2 should read ‘1 but less than 5 miles’.

### Key Points

A general guideline as to when to use these types of questions:

- Use multi-choice questions when you want to count the number of responses in the different categories.
- Use open-ended questions when you want to explore your respondents’ views.
Rating scale

If you’re asking respondents about their attitude towards something or to describe something on an interval scale, then you can use a rating scale. There is some debate over whether you should exclude a neutral category such as neither so that it forces your respondents to make a choice. Examples of rating scales are shown in Example 6.

Example 6

**How satisfied or dissatisfied are you with the local train service?**

*Please circle appropriate number*

- Very satisfied: 1
- Fairly satisfied: 2
- Fairly dissatisfied: 3
- Very dissatisfied: 4
- Not applicable: 5

or

**How satisfied or dissatisfied are you with the local bus service?**

*Please circle appropriate number*

- Very satisfied: 1
- Fairly satisfied: 2
- Neither satisfied nor dissatisfied: 3
- Fairly dissatisfied: 4
- Very dissatisfied: 5

or

**How much pain did you experience after the operation?**

*Please circle appropriate number*

- None: 1
- Very mild: 2
- Mild: 3
- Moderate: 4
- Severe: 5

5.5 Coding

When setting up the questions, you can allocate codes to answers for ease of use for processing the data at a later date. You can code closed questions to a pre-determined scheme; responses to the Other option can be added as common answers emerge.

You should categorise open-ended questions systematically as answers become apparent. We recommend that you put a code-book together to accompany your survey to help those who will eventually process and analyse the data. Usually, the code book would be developed after you have received a significant number of responses to the question, say 200. This should give you sufficient answers to be able to categorise them into groups.

The code-book lists the questions in the questionnaire with details of the response codes; variable names; and labels that will help with the data entry process. There’s an example of a short questionnaire along with coding and the code-book in Appendix 1.
5.6 Routing

Routing should always be used when a certain answer to a question allows you to skip a question that is not relevant. See Example 7.

**Example 7**

**Q1. Do you attend community council meetings?**
*Please circle one number*

- Yes 1 Go to Q3
- No 2 Go to Q2

**Q2. Why do you not attend these meetings?**
*Please circle one number*

- Lack of time 1
- Not interested 2
- Other 3

(please specify) ___________________________

6. Piloting the survey

**Piloting the questionnaire**

It is essential that you test the questionnaire before it is sent out for wider dissemination: this is called a pilot survey. It should be piloted on a smaller sample, usually about 10% of the final sample size or a minimum of about 50 respondents. There are two types of pilot that are useful: *internal* and *external*.

The *internal* pilot is for others in your office to check that the questions are easy to understand and that the right questions get asked in their proper order, (to check that the instructions after each answer lead to the appropriate questions). It is helpful to use people from another department so that you can check to see whether the language you use in the questions is generally understandable and does not contain technical language.

The *external* pilot should be addressed to a sample of individuals who are similar to the population that you are going to survey. Piloting your survey this way will mean that you can do basic checks on the closed-question coding. If too many *Other - please specify* boxes have been completed, you can change the answer categories before the main survey is undertaken to reflect the most common entries. Piloting may be a problem for small populations. In this case, you must find another suitable population to test the pilot questionnaire.

The purpose of a pilot is to check that the questionnaire’s flow and design works in practice. If it doesn’t, the pilot should identify problem questions and get information to help improve the survey’s design. You must re-test significant changes to the questionnaire introduced after the pilot.
Specific areas that should be tested are:

**Questionnaire content and phrasing**
- Are they ambiguous?
- Do you get the information that you need from the questions that have been asked?
- Are questions clear? Receiving responses with missing results will indicate that they are not.

**Layout**
- Is it too cluttered?

**Length**
- Does it take too long to complete? This can be timed during the pilot.

**Pre-coded questions**
- If additional responses are given, these may need to be included in the questionnaire.

Other areas that can be tested from a pilot are:

**Sample size**
- If there’s a large variation in the response to an important question, you may want to consider increasing the sample size.

**Non-response rate**
- If this is high, you’ll need to investigate why.

**Costs**
- Have you accurately assessed the costs?

**Non-response**
You can minimise a survey’s non-response rate by:
- ensuring that the questionnaire is presentable;
- offering incentives - though these do not have to be large (for example, an entry into a free draw for a bicycle in a green transport study);
- notifying respondents about the survey beforehand (this can also give respondents the opportunity to refuse to take part); and
- following up all non-respondents via telephone interviews (where possible).

You should analyse the non-respondents to check for bias i.e. do they come from a similar area or are they from a particular age group? - if that information is available.
7. Analysing the results

7.1 Data entry

You should allocate a unique reference number to each questionnaire so that they can be easily identified during response monitoring and processing: you should do this when you’re compiling the sample. You’ll then need to collate the questionnaires’ responses into a database or spreadsheet.

The database should be set up by defining each column as a question from the questionnaire and then you enter the possible response codes for each of those questions.

For example if you have a question where you can only select one answer option (single code):

**Q1. How long have you lived in this area?**  
*Please circle one number only*

Less than 5 years  1  
5-9 years  2  
10 years or more  3  

There are three possible answers to Q1, but you can only select one, therefore you would enter 1, 2 or 3 in that column in the database.

However, if you had a question where it was possible to select more than one answer option (multi-code):

**Q2. Why have you chosen to live in this area?**  
*Please circle all relevant numbers*

Close to work  1  
Close to friends/family  2  
Area is quiet  3  
Area is affordable  15  

Here, it is possible to select several answers for Q2, therefore every possible answer option will need a separate column in the database. There will be a column for Q2.1 (close to work), Q2.2 (close to friends/family) and so on up to Q2.15 (area is affordable). The responses you put in the database for each of these columns will be either 0 (answer option not selected) or 1 (answer option selected). See Example 8.

---

**Example 8**

<table>
<thead>
<tr>
<th>Unique reference</th>
<th>Q1</th>
<th>Q2.1</th>
<th>Q2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ref2</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ref3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ref4</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Editing the data

Once the data has been entered, you will need to make several types of checks to ensure that there aren’t any errors in the data that’s being analysed. If you miss out this stage of the survey process, you may produce misleading results from the analysis of the data. Any assumptions made about the data should be recorded to support the survey’s findings.

Outliers

Check for out-of-expected-range responses - a full-time worker aged under 16; and data capture errors - the response range is from 1-4 but the respondent (or the person doing the data entry) has entered a 5.

Inconsistencies

Look out for inconsistencies in the data, i.e. the respondent hasn’t followed the routing and has answered questions that they don’t need to. If we look back at Example 8, if the respondent had answered Yes to Q1 but then followed on with explaining why he did not attend the meetings, how should we record this? Can we assume that the respondent meant to say No in Q1? In cases such as these, we need to judge whether to exclude the response or whether to make the whole response logically consistent.

Omissions

Sometimes, we can assume answers to questions where there are missing values. For example, if a respondent did not answer a question on whether they smoked cigarettes but proceeded to answer another on the number of cigarettes smoked per week, we can assume that that they do indeed smoke, and we can fill in the missing answer. If we cannot assume the response, then the data must be coded as Missing (data is usually coded ‘9’, ‘99’ etc., for missing values depending on the number of digits used in the normal answers).

7.2 Analysis

Once you have checked the data and it is error-free, you can carry out the appropriate analysis. You can analyse the data in a number of different ways; the most basic of these ways is through frequency tables and cross tabulations. You should also comment on the survey’s response rate in actual figures and percentages.

Frequency tables

The frequency count is the most common form of analysis for survey data. This gives the number of responses in each category to a particular question and can be additionally shown as proportions or percentages; an example is shown in Table 1. Where there’s some data missing, the percentages in the table can be misleading. That is, Table 1 seems to say that 34% of adults attended the meetings. This, however, ignores the fact that there’s some data missing - the questions were not answered in the questionnaire. Sometimes, you can assume that the data that’s missing is in the same pattern as the data you have. With this data, and taking this assumption into account, the additional column on the right of Table 1a shows that the best estimate of the percentage of adults attending was in fact 38%.
Survey Design & Analysis

Table 1 Attendance of adults at community council meetings

<table>
<thead>
<tr>
<th>Attended?</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>75</td>
<td>34</td>
</tr>
<tr>
<td>No</td>
<td>120</td>
<td>55</td>
</tr>
<tr>
<td>Missing</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>219</td>
<td>100</td>
</tr>
</tbody>
</table>

N=number of respondents

Table 1a Attendance of adults at community council meetings

<table>
<thead>
<tr>
<th>Attended?</th>
<th>N</th>
<th>Per cent with missing category</th>
<th>Per cent without missing category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>75</td>
<td>34</td>
<td>38</td>
</tr>
<tr>
<td>No</td>
<td>120</td>
<td>55</td>
<td>62</td>
</tr>
<tr>
<td>Missing</td>
<td>24</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>219</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Cross tabulations

Cross tabulations are tables that allow you to present the data from one categorical variable broken down by another. This is a useful method as you can compare the question’s response by gender or age group.

Table 2 Attendance of adults at community council meetings by age group

<table>
<thead>
<tr>
<th>Attended?</th>
<th>16-19</th>
<th>20-29</th>
<th>30-39</th>
<th>40-49</th>
<th>50+</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (%)</td>
<td>20</td>
<td>34</td>
<td>48</td>
<td>41</td>
<td>51</td>
<td>38</td>
</tr>
<tr>
<td>No (%)</td>
<td>80</td>
<td>66</td>
<td>52</td>
<td>59</td>
<td>49</td>
<td>62</td>
</tr>
<tr>
<td>Total (%)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Total (N)</td>
<td>46</td>
<td>47</td>
<td>51</td>
<td>42</td>
<td>33</td>
<td>219</td>
</tr>
</tbody>
</table>

N=number of respondents

N.B. This table assumes that those who did not answer have the same attendance record as those that did answer.

From this sample, we could deduce that 38% of the people who were surveyed attend community council meetings. The 95% confidence limits surrounding this estimate are 32% and 44%, so we can be 95% certain that the true percentage of the population attending community council meetings is within this range.

Summary statistics

For responses that are not categorised but are numerical, such as age or earnings, you can produce summary statistics, such as the mean; range; minimum; maximum; and quartiles. This would give you a good indication of the range of data that you’ve received from the survey.

For more information on how to calculate summary statistics please see the guide titled ‘A Guide to Summary Statistics’ on the Data Unit’s website.
However, the main sort of data produced from surveys is categorical: this is best analysed by calculating proportions or percentages as shown in Table 3. Summary statistics such as those described above are meaningless when applied to categorical data since the numbers have no real meaning. Instead, we can look at the total of the satisfied vs the total of the dissatisfied and then the strength of the satisfied etc.

Table 3 Levels of satisfaction with the local bus service

<table>
<thead>
<tr>
<th>Satisfaction (code)</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very satisfied (1)</td>
<td>41</td>
<td>19</td>
</tr>
<tr>
<td>Fairly satisfied (2)</td>
<td>91</td>
<td>41</td>
</tr>
<tr>
<td>Neither satisfied nor dissatisfied (3)</td>
<td>34</td>
<td>16</td>
</tr>
<tr>
<td>Fairly dissatisfied (4)</td>
<td>32</td>
<td>15</td>
</tr>
<tr>
<td>Very dissatisfied (5)</td>
<td>19</td>
<td>9</td>
</tr>
<tr>
<td>Total (N)</td>
<td>217</td>
<td>100</td>
</tr>
</tbody>
</table>

N=number of respondents

8. Reporting

8.1 Presenting the results

You can present survey results via reports, publications, seminars or a website. Your findings, wherever possible, should be fed back to the survey’s respondents. Your report’s content should be tailored to your audience and so you should avoid any over complicated analysis or use of too many figures.

Charts, diagrams and summary tables can help a non-statistician understand the data better than presenting data in large tables. We have prepared another guide called Presenting Data, which aims to help you present messages taken from data in the most appropriate way.

For more information on how to present information please see the guide ‘Presenting Data’, on the Data Unit’s website.
Appendix 1: Example questionnaire

Study of Local Shopping and Transport

As a member of the community, you have been randomly selected to take part in this survey, by completing the enclosed questionnaire.

The aim of this survey is to assess the access you have to local shops and the mode and reliability of the transport that you use. Your response will allow us to determine if improved public transport is needed in your area.

The information from the survey will be treated confidentially and used for statistical purposes only. No personal information will be passed on. The anonymised data will be given to, and used by, Anyshire County Council to determine the needs for developing local shopping and transport in your area.

Please return the completed survey by XX Month 20XX.

By post:
Local Government Data Unit ~ Wales
3-7 Columbus Walk
Cardiff
CF10 4SD

Or by fax: 029 2090 9510

Please answer the following questions as accurately as possible. The information will be treated confidentially.

If you have any queries on this questionnaire, please contact us at the above address, or telephone 029 2090 9500.
Q1. How far is your home from the nearest food shop? (By food shop, we mean somewhere that sells at least groceries e.g. cereals, eggs)

Please circle one number only

Less than 1 mile  1
1 but less than 5 miles  2
5 miles or more  3
Don’t know  4

Q2. How do you normally get to your nearest food shop?

Please circle one number only

Walk  1  go to Q5
Cycle  2  go to Q5
Bus  3  go to Q3
Car  4  go to Q5
Other  5  go to Q5

(Please specify) ________________________________________

Q3. If you travel by bus, in your opinion, how reliable is the bus service?

Please circle one number only

Very reliable  1
Quite reliable  2
Quite unreliable  3
Very unreliable  4

Q4. If you travel by bus, how do you find out information about bus times?

Please circle as many as apply

Word of mouth  1
Traveline  2
Bus driver  3
Other  4

(Please specify) ________________________________________

Personal details
This information is used so that we can be sure that different groups of people are represented in the survey.

Q5. Are you...?

Please circle one number only

Male  1
Female  2

Q6. How old were you on your last birthday?

Please write in whole years

Age (years): 

Thank you for your time, please return your completed questionnaire by XX Month 20XX to:
Local Government Data Unit ~ Wales
3-7 Columbus Walk
Cardiff
CF10 4SD
Or by fax to 029 2090 9510.
**Code – Book**

<table>
<thead>
<tr>
<th>Question</th>
<th>Col</th>
<th>Variable</th>
<th>Variable Label</th>
<th>Code Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>ID</td>
<td>Identifier</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>DISTSHOP</td>
<td>Distance from shop</td>
<td>1</td>
<td>Less than 1 mile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1 but less than 5 miles</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>5 miles or more</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>Don’t know</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>MODETRAN</td>
<td>Mode of transport</td>
<td>1</td>
<td>Walk</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>Cycle</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>Bus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>Car</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>Taxi*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>Lift from colleague*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>Missing</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>MODETRAN</td>
<td>Mode of transport</td>
<td>-</td>
<td>Text given</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>RELBUS</td>
<td>Reliability of bus</td>
<td>1</td>
<td>Very reliable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>Quite reliable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>Quite unreliable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>Very unreliable</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>Missing</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>BUSTIME - WOM</td>
<td>Source of information about bus times - word of mouth</td>
<td>0</td>
<td>Not selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>Word of mouth</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>BUSTIME - TL</td>
<td>Source of information about bus times - traveline</td>
<td>0</td>
<td>Not selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>Traveline</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>BUSTIME - BD</td>
<td>Source of information about bus times - bus driver</td>
<td>0</td>
<td>Not selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>Bus driver</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>BUSTIME - OTHER</td>
<td>Source of information about bus times - other</td>
<td>0</td>
<td>Not selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>Other</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>BUSTIME - WRITE IN</td>
<td>Source of information about bus times - other write in</td>
<td>-</td>
<td>Text given</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>GENDER</td>
<td></td>
<td>1</td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
<td>Missing</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>AGE</td>
<td></td>
<td>-</td>
<td>Digits given</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>98</td>
<td>Aged 98 and over</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>99</td>
<td>Missing</td>
</tr>
</tbody>
</table>

* These will depend on what responses to the ‘Other’ category are received.
Appendix 2: Additional support

The Survey Question Bank

http://surveynet.ac.uk/sqb/about/introduction.asp

The Survey Question Bank (SQB) is co-ordinated by the UK Data Archive (UKDA) at the University of Essex. It provides a suite of online research resources with a specific focus on survey methods and can be used to locate, and view in context, survey questions as they were used in the data collection process.

This website holds the questionnaires for large-scale surveys which have a nationally representative sample and are generally conducted by a large and professional survey organisation.

All the surveys that are held by the SQB are listed by survey name or topic and include:

- an overview of the survey;
- methodology; and
- main topics that have been covered.

The SQB also contains links to relevant datasets.

Harmonised Concepts and Questions


The Harmonised Concepts and Questions booklet, which is available through the Office for National Statistics website is a very good source of questions and associated definitions that are used in government surveys. In addition to the questions, some information is given on the output categories for some variables, such as age, which will then allow comparison of information from one survey to another.
Appendix 3: Further reading

Further Reading


Fink, A. (2003), How to Design Survey Studies, Sage Publications Inc (USA); ISBN: 0761925783


Implications of Data Protection Act 1998 on research and survey; see articles and guidelines on Market Research Society’s website (www.mrs.org.uk/standards/dp.htm)


The Centre for Applied Social Surveys: http://www.s3ri.soton.ac.uk/cass/
